

1 Классификация

1.1 Общий подход к классификации через апостериорные вероятности

Общая подход к классификации: строятся классифицирующие функции f_i , такие что классификация проводится так: индивид с признаками x относится к группе с максимальным значением на нем классифицирующей функции: $\arg \max_i f_i(x)$.

Откуда берутся эти классифицирующие функции? Естественная идея взять в качестве f_i вероятность (ее оценку) принадлежности к i -му классу. Пусть ξ – дискретная с.в., принимающая значения $\{A_i\}_{i=1}^k$, $\mathcal{P}(\eta | \xi = A_i) = \mathcal{P}_i$ и имеет плотность $p_i(x)$. Тогда было бы логично взять $f_i = p_i$. Для практического применения надо было бы оценить плотности, либо непараметрически (например, по числу точек, попавших в дельта-окрестность — типа метода ближайших соседей), либо параметрически (если известно, что распределение нормальное, тогда просто оцениваем векторы средних и ковариационные матрицы).

Более сложный подход — через апостериорные вероятности. Если у нас есть априорное знание вероятности того, что индивид из того или иного класса, то мы можем его учесть. Введем понятие класса $C_i = \{\xi = A_i\}$. Чтобы классифицировать наблюдение x , необходимо найти

$$\arg \max P(\xi \in A_i | \eta = x) = \arg \max P(C_i | x).$$

Пусть известны априорные вероятности принадлежности нового наблюдения к i -му классу $\pi_i = P(C_i)$. Тогда апостериорные вероятности по формуле Байеса будут иметь вид

$$P(C_i | x) = \frac{P(x | C_i) \pi_i}{\sum_{j=1}^k P(x | C_j) \pi_j}.$$

Поэтому в качестве классифицирующих функций берут

$$f_i(x) = \frac{p_i(x) \pi_i}{\sum_{j=1}^k p_j(x) \pi_j}.$$

Так как знаменатель у всех f_i одинаковый, его можно отбросить, и итоговые классифицирующие функции будут выглядеть как $f_i(x) = P(x | C_i) \pi_i = p_i(x) \pi_i$.

Как выбрать априорные вероятности?

1. Равномерно, $\forall i \in 1 : k \pi_i = 1 / k$.
2. По соотношениям в обучающей выборке: $\pi_i = n_i / \sum_{j=1}^k n_j$.
3. На основе другой дополнительной информации о данных (результаты предыдущих исследований, etc.)

Свойство. Построенный метод классификации $\text{predict}(x) = \arg \max_i \pi_i p_i(x)$ минимизирует среднюю апостериорную ошибку:

$$\sum_{i=1}^k \pi_i P(\text{predict}(x) \neq i | C_i).$$

Видно, что можно с помощью априорных вероятностей формально задавать важность ошибочных классификаций для разных классов.